NATIONAL FORUM JOURNAL OF COUNSELING AND ADDICTION
VOLUME 2, NUMBER 1, 2013

The Penn State Worry Questionnaire: Comparative Model Fit and Invariance across Gender

Janet A. Carter, PhD
Assistant Professor
College of Education
Department of Educational Leadership, Foundations, and Counseling
The University of Texas of the Permian Basin
Odessa, TX

Scott C. Bates, PhD
Associate Professor
Emma Eccles Jones College of Education and Human Services
Department of Psychology
Utah State University
Logan, UT

Abstract

This article identified *a priori* models of the Penn State Worry Questionnaire (PSWQ) identified in the extant literature and utilized bootstrapped confirmatory factor analysis (CFA) to investigate the comparative fit of the models in a nonclinical sample of college students (N = 502). Results indicated that CFA models with method effects provided relatively better fit in comparison to the original one-factor model proposed by Meyer et al. (1990). An abbreviated, 8-item model provided a significant improvement in fit compared to 16-item models, and the present study extended assessed measurement invariance across gender for the abbreviated model. Results indicated that the abbreviated model displayed configural and factor loading invariance across gender but did not maintain scalar equivalence across gender. Results also indicated significant differences in terms of latent mean structure, with women displaying significantly higher levels of latent worry than men. Implications of scalar invariance are discussed.

Keywords: worry, Penn State Worry Questionnaire, structural equation modeling, comparative fit, measurement invariance

Thought processes influence individual behavior and emotions, and psychological research has specifically focused on the role of personal appraisals and cognitive processes associated with the onset, maintenance, and duration of psychological disorders. One such cognitive process, worry, initially garnered research interest in relation to test anxiety (Deffenbacher, 1980; Liebert & Morris, 1967; Morris Davis, & Hutchings, 1981) and in connection to insomnia (Borkovec, Robinson, Pruzinsky, & DePree, 1983). Academic and clinical interest in the worry construct increased following the publication of the revised third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-III-R*; American Psychiatric Association [APA], 1987), which included generalized anxiety disorder (GAD) as a separate diagnostic category and identified chronic, uncontrolled worry as its main diagnostic criterion (Borkovec, 1994; Craske, 1999; Holaway, Rodebaugh, & Heimburg, 2006). Worry remained the core diagnostic feature of GAD in the succeeding editions of the *DSM* (1994, 2000, 2013; APA), and subsequent research implicated worry in other anxiety disorders such as panic disorder, obsessive-compulsive disorder, post-traumatic stress disorder, and social phobia (Andrews & Borkovec, 1988; Barlow, 2002; Borkovec et al., 1983; Craske, Rapee, Jackel, & Barlow, 1989; Papageorgiou, 2006; Tallis & de Silva, 1992; Watts, Coyle, & East, 1994).

As interest in the worry construct increased, Meyer, Miller, Metzger, and Borkovec (1990) observed that anxiety research primarily assessed worry by asking individuals what percentage of time they typically spent worrying. Meyer and his colleagues noted that existing assessment instruments focused on worry versus emotionality or cognitive versus somatic complaints. They recognized the need for a psychometrically sound self-report measure of trait worry and subsequently developed the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990). Meyer et al. initially conceptualized the PSWQ as a unifactorial measure, with 16 Likert-scale items contributing to a general worry factor that reflected frequency and intensity of worry without regard to specific content. The PSWQ has well-established psychometric properties (Beck, Stanley, & Zebb, 1995; Borkovec, 1994; Brown, Antony, & Barlow, 1992; Davey, 1993; Meyer et al., 1990; Molina & Borkovec, 1994; Pallesen, Nordhus, Carlstedt, Thayer, & Johnsen, 2006; Stöber, 1998; van Rijsoort, Emmelkamp, & Vervaeke, 1999) and now represents the predominant worry measure utilized in the extant literature.

In spite of strong psychometric properties, considerable debate has revolved around the latent factor structure of the PSWQ. Both Meyer et al. (1990) and Brown et al. (1992) retained a one-factor solution based on principal component analysis (PCA) and the resulting eigenvalues and scree plots obtained from nonclinical and clinical samples, respectively. Fortune, Richards, Griffiths, and Main (2005) also retained a one-factor solution following exploratory factor analysis (EFA) with a medical sample. In contrast, results from other factor analytic studies (Beck et al., 1995; Fresco, Frankel, Mennin, Turk, & Heimberg, 2002; van Rijsoort et al., 1999; Yilmiz, Gencöz, & Wells, 2008) have indicated that the PSWQ includes two factors: one factor including the 11 positively-worded items (2, 4, 5, 6, 7, 9, 12, 13, 14, 15, and 16) and a separate but correlated factor including the 5 reverse-scored items (1, 3, 8, 10, and 11). Several CFA studies have compared the one- and two-factor models but results have been mixed, with some studies finding general support for the two-factor model (Castillo, Macrini, Cheniaux, & Landeira-Fernandez, 2010; Fresco, Heimberg, Mennin, & Turk, 2002; Hazlett-Stevens, Ullman, & Craske, 2004; Olatunji, Schottenbauer, Rodriquez, Glass, & Arnkoff, 2007) and other CFA studies suggesting that the two-factor model provided inadequate fit for the data (Brown, 2003; Carter et al., 2005; Castillo et al., 2010; Gana, Martin, Canouet, Trouillet, & Meloni, 2002; Hopko et al., 2003; Pallesen et al., 2006).

In light of the mixed findings regarding the factor structure of the 16-item PSWQ, Brown (2003) presented a cogent argument that the two-factor model resulted from method effects introduced by the positively- and negatively-worded (reverse scored) items (Marsh, 1996). He subsequently tested a model that included a general worry factor as well as error covariance among the five negatively-worded items and between positively-worded items that displayed content overlap (Items 7 and 15 and Items 9 and 16). Researchers have also accounted for method effects by specifying one general worry factor consisting of all 16 items and a method-effect factor consisting of the five negatively-worded items (Hazlett-Stevens et al., 2004; van der Heiden, Muris, Bos, & van der Molen, 2010). Other studies have incorporated method effects by including one general worry factor and two orthogonal factors: one factor that accounted for method effects among the positively-worded items and one that included method effects among the negatively-worded items (Castillo et al., 2010; Gana et al., 2002; Hazlett-Stevens et al., 2004; Pallesen et al., 2006). However, the extent of method effects in the 16-item PSWQ is not yet fully understood because the CFA studies have involved foreign-language translations of the PSWQ; younger and older participants; and student, community, and clinical samples. Further research investigating the factor structure of the PSWQ appears warranted.

Given the concerns about methods effects in the 16-item PSWQ, researchers (Brown, 2003; Castillo et al., 2010; Fortune et al., 2005; Fresco, Heimberg et al., 2002; Gana et al., 2002; van der Heiden et al., 2010; van Rijsoort et al., 1999) have questioned whether the negatively-worded items in the two-factor model, which is frequently interpreted as an absence-of-worry factor, represents a substantive or theoretically meaningful dimension of trait worry. Doubts about the conceptual veracity of the absence-of-worry factor are further compounded by evidence that the negatively-worded items function differently within and across samples. For instance, Brown et al. (1992) utilized a clinical sample of individuals diagnosed with anxiety disorders, and found that omission of Item 1 (“If I do not have enough time to do everything, I do not worry about it”), the item with the lowest factor loadings in the Meyer et al. (1990) study, improved reliability estimates across diagnostic groups. Van Rijsoort et al. (1999) also noted that Item 1 displayed the poorest corrected item-total correlation in their community sample. Pallesen et al. (2006) developed a Norwegian adaptation of the PSWQ and found that Items 1 and 11 (“When there is nothing more I can do about a concern, I do not worry about it any more”) produced a greater than 5% drop in explained variance within their community sample in comparison to the amount of explained variance in their student sample. When comparing EFA models across White-American and African-American college student samples, Carter et al. (2005) found that the factor including the positively-worded items from the PSWQ demonstrated equivalence across groups, but Item 10 (“I never worry about anything”) loaded on an absence-of-worry factor comprised of the 5 negatively-worded items for the White-American sample. For the African American sample, Item 10 loaded on a third worry dismissal factor. Fresco, Frankel et al. (2002) found that Item 11 loaded across multiple factors. Aside from the apparent inconsistencies in how the negatively-worded items cohere within an absence-of-worry factor, other research has demonstrated that the factor typically contributes very little to explained variance in comparison to the general worry factor (Castillo et al., 2010; Olantunji et al., 2007).

In keeping with concerns about method effects and lack of cohesion among the negatively-worded items, Brown (2003) suggested that the 11 positively-worded items potentially represent an adequate and more theoretically sound measure of trait worry, and he called for further CFA studies to test the feasibility of such a measure. Hazlett-Stevens et al. (2004) modeled paths from worry and a method factor comprised of the negative items and found that removal of the method factor did not significantly degrade the model. In light of these findings, it is important to examine abbreviated, one-factor models to determine if they provide a better representation of trait worry. No study to date has specifically used CFA to examine the comparative fit of an abbreviated model that totally eliminated the negatively-worded items included in the PSWQ, so this study examines a one-factor, 11-item CFA model. In addition, this study examines another abbreviated model of the PSWQ presented by Hopko et al. (2003) and Crittendon and Hopko (2006). Hopko et al. (2003) used EFA to examine the factor structure of the PSWQ in a clinical sample of older adults with GAD and noted that neither the single-factor nor two-factor CFA models provided adequate fit for the data. Through a series of model specification procedures, they eliminated all of the reverse-scored items (1, 3, 8, 10, 11) and three positively-worded items (14, 15, 16), creating an abbreviated, 8-item measure referred to as the PSWQ-A. The elimination of Items 15 and 16 fit with Brown's (2003) findings that the two items significantly contributed to method effects related to content overlap. Crittendon and Hopko (2006) again used EFA and found that the factor structure of the PSWQ-A held across a community sample of older adults and a college-student sample. Like the 11-item model, this one-factor, 8-item model has not yet been examined using CFA.

Although there is a growing body of CFA research that indicates that the PSWQ (Meyer et al., 1990) includes method effects, further investigation is warranted for two reasons. First, studies with English-speaking samples have been largely limited to comparisons between the 16-item, one-factor model and the 16-item, two-factor model with separate but correlated factors for the positively- and negatively-worded items that represented worry and absence of worry (Carter et al., 2005; Olatunji et al., 2007). However, the extant literature includes several competing models that incorporate method effects and that could provide viable alternatives to the two-factor model. Only Hazlett-Stevens et al. (2004) compared additional CFA models, but the results indicated that the model with a general worry factor and two orthogonal method-effect factors did not provide good fit for the data because all but three positively-worded items did not significantly load on the positive method-effects factor, and these results were replicated in a second sample. Second, the CFA research that has documented the viable alternative CFA models has been produced with foreign-language translations of the PSWQ (Castillo et al., 2010; Gana et al., 2002; Pallesen et al., 2006; van der Heiden et al., 2010), and replication in an English-speaking sample seems like a logical next step in moving toward consensus regarding the factor structure of the PSWQ. Accordingly, the first part of this study examined the comparative fit of five different CFA models presented in the literature: 1) the 16-item, one-factor model originally conceptualized by Meyer et al. (1990), 2) the 16-item, one-factor model proposed by Brown (2003) with correlated errors among select positively-worded items (7 and 15, 9 and 16) and all of the negatively-worded items, 3) the 16-item, two-factor model that included separate correlated factors consisting of positively-worded items and negatively-worded items (Carter et al., 2005; Fresco, Frankel et al., 2002; Fresco, Heimberg et al., 2002; Hopko et al., 2003; Olatunji et al., 2007), 4) a two-factor model that included one general worry factor and a method effect factor the included negatively-worded items, and 5) a three-factor model that included a general worry factor and two orthogonal method effects factors. Given Brown's (2003) assertion that the two-factor model allowed for error variance associated with negatively-worded items, it was hypothesized that the two-factor solution would provide better fit than the 16-item, one-factor model.

In addition to the five nested model comparisons for the 16-item PSWQ, this study assessed two abbreviated models: an 11-item, one-factor model consisting of all positively-worded items (2, 4, 5, 6, 7, 9, 12, 13, 14, 15, and 16) from the PSWQ and an abbreviated, one-factor model that included the eight positively-worded items (2, 4, 5, 6, 7, 9, 12, and 13) retained by Hopko and his colleagues (PSWQ-A; Crittendon & Hopko, 2006; Hopko et al., 2003). Since the abbreviated models eliminated the negatively-worded items that have demonstrated inconsistency in past studies, it was that the abbreviated models would provide better fit than the 16-item, one- and two-factor models, but no specific hypotheses were made regarding the comparative fit between the 8-item and 11-item models.

Along with examining the comparative fit of the various PSWQ models, this study examined configural, measurement, and latent mean invariance across gender using CFA. Although multiple studies have examined the factor structure of the PSWQ, only Brown (2003) and Pallesen et al. (2006) have previously examined invariance across gender. Brown utilized a clinical sample of individuals with anxiety disorders and tested invariance across gender in the 16-item, one-factor model that included correlated errors among the negatively-worded items and between Items 7 and 15 and between Items 9 and 16 (Model 2). He found evidence of configural and measurement invariance across split samples and also found a significant latent mean difference in terms of the worry variable, with women displaying significantly higher means than men. Pellesen et al. (2006) utilized a Norwegian community sample and examined invariance in the 16-item, three-factor model that included a general factor and two orthogonal method-effect factors that represented the positively- and negatively-worded items (Model 4). Results of the invariance tests indicated significant latent mean differences on the general worry variable and the reverse-scored items, with women scoring higher than men on both variables. In contrast, men demonstrated higher latent mean differences on the positively-worded variable. Pellesen et al. also found variance across gender in terms of latent variable variances and error variances. Although the samples in these studies differed in terms of language and type (clinical vs. community), these findings suggest that method effects could potentially cloud real latent mean differences between men and women, so invariance across gender needs to be included in the CFA process to verify that the PSWQ items operate similarly across groups. This study seeks to replicate previous research by examining invariance across gender and adds to the existing literature by examining invariance in the abbreviated PSWQ models that have not yet been subjected to CFA.

Method

Participants

Two independent samples of college students were recruited from undergraduate and graduate psychology courses at a large western university in consecutive fall (n = 260) and spring (n = 305) semesters. Participants completed a questionnaire packet that included a demographic information form and a set of self-report measures. Completed questionnaires were screened for missing data, and packets were removed from the data set if questionnaires were incomplete. A total of 63 questionnaire packets were excluded from data analysis due to missing information (fall = 16, spring = 47). The two samples were combined following data screening since the two groups did not differ in terms of sex ($\chi^2_{(1,\ 502)} = 3.03$, $p = .08$), ethnicity ($\chi^2_{(5,\ 502)} = 8.81$, $p = .12$), or religious affiliation ($\chi^2_{(5,\ 497)} = 2.31$, $p = .81$). The demographic characteristics of the sample are presented in Table 1.

Table 1

Demographic Characteristics

	Fall		Spring		Combined	
	N	%	*N*	%	*N*	%
Gender						
Male	81	(33.2)	105	(40.7)	186	(37.1)
Female	163	(66.8)	153	(59.3)	316	(62.9)
Ethnicity						
European American/White	224	(91.8)	224	(86.8)	448	(89.2)
African American			5	(1.0)	5	(1.0)
Hispanic/Latino	11	(4.5)	9	(3.5)	20	(4.0)
Asian or Asian American	4	(1.6)	5	(1.9)	9	(1.8)
Native American			3	(1.2)	3	(0.6)
Other	5	(2.0)	8	(3.1)	13	(2.6)
Not Specified			4	(1.6)	4	(0.8)
Marital Status						
Single	192	(78.7)	232	(89.9)	424	(84.5)
Married	45	(18.4)	16	(6.2)	61	(12.2)
Divorced, Not Remarried	6	(2.5)	6	(2.3)	12	(2.4)
Widowed	1	(0.4)			1	(0.2)
Not Specified			4	(1.6)	4	(0.8)
Education Level						
Freshman	78	(32.0)	152	(58.9)	230	(45.8)
Sophomore	48	(19.7)	65	(25.2)	113	(22.5)
Junior	65	(26.6)	24	(9.3)	89	(17.7)
Senior	45	(18.4)	11	(4.3)	56	(11.2)
Graduate Student	7	(2.9)	1	(0.4)	8	(1.6)
Not Specified	1	(0.4)	5	(2.0)	6	(1.2)
Religious Affiliation						
Protestant	7	(2.9)	11	(4.3)	18	(3.6)
Catholic	5	(2.0)	9	(3.5)	14	(2.8)
LDS (Mormon)	210	(86.1)	208	(80.6)	418	(83.3)
Buddhist	1	(0.4)	2	(0.2)	3	(0.6)
Other	20	(8.2)	22	(8.5)	42	(8.4)
Not Specified	1	(0.4)	6	(2.3)	7	(1.4)

Data Collection Procedures

The researcher recruited participants from undergraduate and graduate courses. All participants were entered in a random drawing for music or book gift certificates in exchange for participation, and a portion of the students also received extra credit from course instructors. Participants attended a single research session, and the researcher introduced the study as an investigation of the relationship between an individual's thought processes and reported feelings of anxiety and/or depression. After providing informed consent, all participants completed a questionnaire packet that included a demographic information form followed by the research measures. The order of the administration of the research measures was counterbalanced across participants.

Measures

Demographic information. Participants completed a brief background questionnaire constructed for this study that detailed age, sex, ethnic background, marital status, current education level, and religious affiliation.

Penn State Worry Questionnaire (PSWQ). The PSWQ is a 16-item self-report questionnaire designed to assess the frequency and intensity of general worry. Each item is rated on a scale of 1 ("not at all typical of me") to 5 ("very typical of me"). Possible scores range from 16 to 80, with high scores representing greater reported levels of general worry. Five of the questionnaire items (1, 3, 8, 10, and 11) are reverse scored to minimize the effects of acquiescence.

The PSWQ demonstrated good psychometric properties in Meyer et al.'s (1990) initial validation studies, with reported alpha coefficients ranging from .91 to .95 in four separate studies. Subsequent research across nonclinical, clinical, and community samples produced reliability estimates ranging from .86 to .95 (Borkovec, 1994; Brown et al., 1992; Davey, 1993; van Rijsoort et al., 1999). The PSWQ also demonstrated adequate test-retest reliability estimates over 2 to 10 weeks (r = .74 - .92) in college samples (Meyer et al., 1990; Molina & Borkovec, 1994; Stöber, 1998). Meyer et al. (1990) reported evidence of validity for the PSWQ as a measure for assessing trait worry. The Chronbach's coefficient alpha for the current sample was 0.94.

Beck Depression Inventory (BDI). The BDI (Beck, Ward, Mendelson, Mock, & Enbaugh, 1961) is a 21-item questionnaire that assesses the affective, cognitive, behavioral, and somatic symptoms of depression. Each item includes four statements that respondents rate from 0 to 3 in terms of perceived intensity. In a meta-analysis of the psychometric properties of the BDI, Beck, Steer, and Garbin (1988) reported evidence of concurrent validity from a total of 35 studies that included correlations between the BDI and other depression measures. Additionally, Beck et al. (1988) summarized multiple studies that support the construct validity of the BDI. In terms of reliability, they reported an average internal consistency coefficient of .86 for psychiatric patients and .81 for nonpsychiatric patients. Cronbach's alpha for this sample equaled .90.

Beck Anxiety Inventory (BAI). The BAI (Beck, Epstein, Brown, & Steer, 1988) is a 21-item questionnaire that assesses the severity of anxiety. Beck et al. derived the BAI items from three self-report measures that assessed anxiety (Beck & Steer, 1993). Respondents rate descriptive statements of anxiety symptoms on a 4-point scale ranging from 0 ("Not at all") to 3 ("Severely, I could barely stand it"). The maximum score is 63 points. Beck and Steer (1993) noted that content validity of the BAI was addressed by inclusion of symptoms representative of anxiety, especially those for panic disorder and GAD. Additionally, Beck and Steer report ample support for convergent, construct, discriminant, and factorial validity. Initial validations studies (Beck et al., 1988) reported high internal consistency estimates, with coefficient alphas of .92 in a sample of outpatients and .94 in a clinical sample. Chronbach alphas for other clinical samples ranged from .85 to .93 (Beck & Steer, 1993). Test-retest correlations after one week equaled .75, $p < .001$ (Beck & Steer, 1993). Cronbach's alpha for this sample equaled .91.

Data Analysis

Pre-analysis data screening. After elimination of cases with missing data, preparation for structural equation model testing in the Analysis of Moment Structures (AMOS 18) involved review of the inter-item correlation matrices to ensure that all summative items displayed positive correlations. Additionally, data screening included examination of linear regression scatterplots (with case numbers entered as the dependent variable) and normal and detrended probability plots. Examination of skewness and kurtosis statistics supplemented visual inspection of the plots. Univariate normality in the sample was assessed through examination of the absolute skewness and kurtosis standardized values, with skewness values greater than 3.00 (Kline, 2005) and kurtosis values equal to or greater than 7 (Byrne, 2010) suggesting univariate nonnormality. Additionally, multivariate kurtosis was assessed by examining Mardia's normalized estimate of multivariate kurtosis (the multivariate *Z*-value in AMOS), with values greater than 5.00 suggesting multivariate nonnormality (Byrne, 2010).

Model testing. Because of the apparent multivariate kurtosis, CFA models of the PSWQ (Meyer et al., 1990) were estimated in AMOS 18 using Bollen-Stine bootstrap procedures with Maximum Likelihood (ML) estimation. Subsequent evaluation of model fit utilized the Bollen-Stine bootstrap *p* value and bootstrap adjusted chi-square and goodness-of-fit statistics. Assessment of goodness of fit for the models involved multiple fit indexes including the chi-square (χ^2) statistic with the Bollen-Stine bootstrap *p* value and the χ^2 likelihood ratio statistic (shown as CMIN/DF in AMOS). Schumacker and Lomax (2004) noted that χ^2 likelihood values of 5 or less indicated adequate fit, while Garson (2008) recommended values of less than 3 but more than 1 as indicators of good model fit.

Based on research precedents (Byrne, Stewart, Kennard, & Lee, 2007) and Blunch (2008), CFA models were also evaluated for goodness of fit based on the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1993), and the Standardized Root Mean Square Residual (SRMR). CFI and TLI values of .95 or more suggest good fit (Byrne, 2010; Byrne et al., 2007). RMSEA has several advantages as a measure of fit because it approximates a noncentral chi-square distribution, which does not require a true null hypothesis or perfect fit in the population (Kline, 2005), and it includes a correction for model complexity and sample size. By convention, models with RMSEA values between .06 to .08 represent adequate fit and models with values less than or equal to .05 represent good fit (Shumacker & Lomax, 2004). Models with RMSEA values greater than .10 should not be accepted (Blunch, 2008). Confidence intervals (90% level) for the RMSEA values are included with the measures of fit, with interval values including.05 indicative of adequate fit and interval values less than .05 indicative of good fit (MacCallum, Browne, & Sugawara, 1996). The closeness-of-fit statistic (PCLOSE) tests the null that the population

RMSEA is no greater than .05. If PCLOSE is less than .05, one may reject the null and conclude that the computed RMSEA is greater than .05 and indicative of poor fit (Garson, 2008). Jöreskog and Sörbom (as cited in Byrne, 2010) recommended that the PCLOSE value exceed .50. The SRMR reflects the overall difference between observed and predicted correlations derived from the covariance matrixes (Kline, 2005). SRMR values of .10 or less indicate adequate fit (Kline, 2005), although Hu and Bentler (1999) suggested values close to .08. The Akaike Information Criterion (AIC) and the Consistent Akaike Information Criterion (CAIC) were included to assess model parsimony (Lee et al., 2008), and the and Expected Cross-Validation Index (ECVI) values were included when alternative, non-nested models were compared. Lower values indicate a higher probability of model replication. Table 2 summarizes criteria for evaluating goodness of model fit and cut points for differentiating between adequate and good models.

Table 2

Criteria for Evaluating CFA Model Fit

	Model Fit	
Criterion	Adequate	Good
CMIN/DF	≤ 5	≤ 3
CFI	≥ .90	≥ .95
TLI	≥ .90	≥ .95
RMSEA	≤ .08	≤ .05
RMSEA Confidence Intervals	Interval includes .05	Interval values < .05
SRMR	≤ .10	≤ .08
PCLOSE		> .05

CMIN/DF = χ^2 likelihood ratio statistic; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual; PCLOSE = *p* value for testing close fit

Results

Descriptive Statistics

The means, standard deviations, skewness, and kurtosis values for the 16 items comprising the PSWQ (Meyer et al., 1990) are reported in Table 3. The results indicated univariate normality in terms of skewness and kurtosis, but Mardia's normalized estimate of multivariate kurtosis (*Z* = 22.42) indicated that the PSWQ did not meet the underlying assumption of multivariate normality.

Table 3

Means, Standard Deviations, Skewness, Kurtosis, and Z-values for the PSWQ

Variable	Mean	SD	Skew	Z-value	Kurtosis	*Z*-value
Item 1	3.49	1.08	-.54	-.4.92	-.43	-1.98
Item 2	2.24	1.01	.78	7.09	.29	1.35
Item 3	3.35	1.16	-.29	-2.62	-.90	-4.13
Item 4	2.29	1.03	.57	5.20	-.25	-1.16
Item 5	2.32	1.18	.62	5.67	-.54	-2.47
Item 6	2.91	1.12	.24	2.17	-.66	-3.01
Item 7	2.13	1.12	.90	8.21	.16	.75
Item 8	3.12	1.10	-.02	-.13	-.83	-3.79
Item 9	2.43	1.07	.62	5.62	-.12	-.56
Item 10	3.92	1.09	-.59	-5.42	-.81	-3.69
Item 11	2.93	1.16	.00	.03	-.94	-4.28
Item 12	2.29	1.24	.72	6.58	-.48	-2.18
Item 13	2.67	1.19	.36	3.25	-.70	-3.21
Item 14	1.96	0.99	.97	8.91	.56	2.56
Item 15	1.91	1.05	1.16	10.57	.77	3.51
Item 16	2.95	1.15	.25	2.32	-.61	-2.79
Multivariate					48.04	22.42

Confirmatory Factor Analyses

Since researchers have not yet reached consensus regarding the factor structure of the PSWQ, this study examined the comparative fit of five different models presented in the literature: 1) the 16-item, one-factor model (Meyer et al., 1990), 2) Brown's (2003) 16-item, one-factor model that allowed for residual covariation among the negatively-worded items and select positively-worded items, 3) the 16-item, two-factor model that included separate but correlated factors for positively- and negatively-worded (Carter et al., 2005; Fresco, Frankel et al., 2002; Fresco, Heimberg et al., 2002; Hopko et al., 2003; Olatunji et al., 2007), 4) a two-factor model with a general worry factor and a method-effect factor that included the five negatively-worded items, and 5) a 16-item, three-factor model that included a general worry factor and two orthogonal method factors. This study also examined an 11-item, one-factor model consisting of all positively-worded items and the 8-item PSWQ-A model proposed by Hopko and his colleagues (Crittendon & Hopko, 2006; Hopko et al., 2003).

Comparative fit of the 16-item PSWQ models. The initial model comparisons involved examination of the nested 16-item models (Models 1, 2, 3, 4, and 5), and the goodness-of-fit statistics for these models are provided in Table 4.

Table 4

Comparisons and Goodness-of-Fit Statistics for the Penn State Worry Questionnaire CFA Models

		B-S χ^2	*df*	ΔB-S χ^2	Δdf	CMIN/ DF	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR	PCLOSE	AIC	CAIC
<u>16-item nested models</u>														
1)	One-factor	510.81	104			4.91	.92	.91	.09	(.08 - .10)	.05	.00	574.81	741.81
2)	One-factor with error	284.34	92	226.47***	12	3.09	.96	.95	.07	(.06 - .07)	.03	.00	372.34	601.96
3)	Two-factor	366.85	103	143.96***	1	3.56	.95	.94	.07	(.06 - .08)	.04	.00	432.85	605.06
4)	Worry with one method factor	358.41	99	152.40***	5	3.62	.95	.94	.08	(.06 - .08)	.04	.00	432.41	625.50
5)	Worry with two method factors	275.30	88	235.51***	16	3.13	.96	.95	.07	(.06 - .07)	.03	.00	371.30	621.79

Note: B-S χ^2 = Bollen-Stine bootstrap adjusted chi-square (all *p*s < .01); CMIN/DF = χ^2 likelihood ratio statistic; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual; PCLOSE = *p* value for testing close fit; AIC = Akaike Information Criterion; CAIC = Consistent Akaike Information Criterion
*** *p* < .001

Model 1, the unidimensional model originally proposed by Meyer et al. (1990), served as a baseline model for nested model comparisons with the four other models derived from the literature (Models 2, 3, 4, and 5). The results indicated that Model 1 displayed the lowest CFI and TLI values and the highest RMSEA values, which represented relatively poor fit for the data in comparison to the other models that incorporated error covariance and method effects (Models 2 - 5). As hypothesized, the two-factor model (Model 3) that included separate but correlated worry and absence-of-worry factors provided better fit for the data than did Model 1, but all models that incorporated method effects produced statistically significant improvements in fit compared to Model 1. Brown’s (2003) correlated error model (Model 2) and the three-factor model that included two orthogonal method factors (Model 5) represented the best-fitting models when comparing the fit indexes across all models, and these models produced the largest changes in CFI (.04 for both) in comparisons with Model 1. However, the factor that represented the method effects associated with the positively-worded items in Model 5 does not appear conceptually valid in that only two of the regression estimates (Items 6 and 9) were statistically significant. Of the remaining four models, Brown’s (2003) one-factor model with correlated error (Model 2) provided the best fit among the nested models. However, the goodness-of-fit statistics did not meet the criterion cutoffs indicative of good fit established for this study in that the χ^2 likelihood ratio exceeded 3.0, RMSEA values exceeded .05, RMSEA confidence interval values exceeded .05, and PCLOSE values did not reach .05. These results suggested that Brown’s model still contained possible points of misfit.

Comparative fit of the abbreviated PSWQ models. Examination of the goodness-of-fit statistics for the two abbreviated PSWQ models indicated that the 11-item, one-factor model comprised of all the positively-worded from the PSWQ provided adequate rather than good fit for the data (B-S $\chi^2_{(44)}$ = 220.70, Bollen-Stine bootstrap p = .00; CMIN/DF = 5.02, CFI = .95, TLI = .94, RMSEA = .09, CI = .08 - .10, SRMR = .04, PCLOSE = .00). Only the 8-item PSWQ-A presented by Hopko et al. (2003) and Crittendon and Hopko (2006) displayed good fit for the data according to the criterion cutoffs established for this study (B-S $\chi^2_{(20)}$ = 59.43, Bollen-Stine bootstrap p = .01; CMIN/DF = 2.97, CFI = .98, TLI = .98, RMSEA = .06, CI = .05 - .08, SRMR = .03, PCLOSE = .12). The 8-item PSWQ-A model clearly represented the best fit among the seven tested models included in this study. However, the RMSEA value, the associated confidence interval values, and the PCLOSE statistic indicated that this model, like the others, still contained possible points of misfit.

Model Modification

Although the PSWQ-A (Crittendon & Hopko, 2006; Hopko et al., 2003) provided the best fit among the different models, the fit statistics for the abbreviated scale suggested that the model still potentially contained points of strain. The modification indexes and the standardized residual covariances for the 8 items in the PSWQ-A were examined to determine whether model respecification would improve fit. The modification statistics associated with the regression weights and covariances indicated possible content overlap between Items 12 (“I have been a worrier all my life”) and 13 (“I notice that I have been worrying about things”), and the highest standardized residual covariance (2.15) occurred between these same items. Freely estimating the error covariance between Items 12 and 13 resulted in a statistically significant improvement in model fit ($\Delta\chi^2_{(1)}$ = 22.00, p < .001), and the modified model displayed good fit statistics for the entire sample ($\chi^2_{(19)}$ = 37.43, Bollen-Stine bootstrap p = .11 (*ns*); CMIN/DF = 1.97, CFI = .99, TLI = .99, RMSEA = .04, CI = .02 - .07, SRMR = .02, PCLOSE = .66). Model modification appeared justified given the redundancy between Items 12 and 13, the improvement in goodness-of-fit statistics, and the statistical significance of the added error covariance parameter (p < .001). The respecified model that included the added error parameter, the modified PSWQ-A, was retained for subsequent testing for invariance across gender.

Prior to testing for invariance across gender, separate models were conducted to test adequate fit for men (n = 186) and women (n = 316). The results indicated good fit for men ($\chi^2_{(19)}$ = 21.00, Bollen-Stine bootstrap p = .72 (*ns*); CMIN/DF = 1.11, CFI = 1.00, TLI = 1.00, RMSEA = .02, CI = .00 - .70, SRMR = .03, PCLOSE = .78) and good fit for women ($\chi^2_{(19)}$ = 31.63, Bollen-Stine bootstrap p = .22 (*ns*); CMIN/DF = 1.67, CFI = .99, TLI = 1.00, RMSEA = .05, CI = .01 - .07, SRMR = .02, PCLOSE = .56). The factor loadings and factor determinants for men, women, and the total sample are presented in Table 5. Examination of correlations between the full PSWQ and the modified PSWQ-A indicated that scores from the 8-item model were highly correlated with scores from the 16-item PSWQ (total sample r = .97, men r = .96, and women r = .97, all ps < .01), suggesting that the modified PSWQ-A still represented the underlying worry construct tapped by the full-length measure developed by Meyer et al. (1990).

Table 5

Latent Structure of the PSWQ-A

Item		Total Sample	Men	Women
2	My worries overwhelm me	.77	.69	.79
4	Many situations make me worry	.84	.83	.84
5	I know I shouldn't worry about things, but I just cannot help it	.88	.84	.90
6	When I am under pressure I worry a lot	.74	.74	.73
7	I am always worrying about something	.80	.72	.84
9	As soon as I finish one task, I start to worry about everything else I have to do	.68	.59	.70
12	I've been a worrier all my life	.59	.57	.59
13	I notice that I have been worrying about things	.72	.76	.70
	Factor determinants	.05	.03	.05
	Scale reliability	.91	.90	.92
	Mean	19.28	17.52	20.32
	SD	7.07	6.35	7.28

Note: All standardized regression weights significant (p < .001)

Tests of invariance across gender for the modified PSWQ-A. Brown (2003) and Pellesen et al. (2006) conducted CFA to examine measurement invariance across gender in the full 16-item PSWQ (Meyer et al., 1990). Invariance across gender has not yet been assessed in abbreviated models (Crittendon & Hopko, 2006; Hopko et al., 2003), so this study examined possible sources of variance across gender in the modified PSWQ-A. Following guidelines outlined by Byrne and her colleagues (Byrne, 2010; Byrne, Shavelson, & Muthén, 1989), invariance testing included sequential examination of configural invariance (reference model), factor loadings invariance, factor variance-covariance invariance, and item uniqueness or error invariance. Additionally, this study examined the invariance of the latent mean structure of the modified PSWQ-A across gender following procedures outlined by Byrne et al. (1989), which involved sequential examination of factor loadings invariance (reference model), item intercepts (scalar) invariance, and latent means invariance. Accordingly, invariance across gender was first tested by conducting a simultaneous analysis of data from both samples of men and women to evaluate the equality of the modified model between sexes. The results indicated that the simultaneous reference model provided a good fit for the data ($\chi^2_{(38)}$ = 52.63, Bollen-Stine bootstrap p = .46; CMIN/DF = 1.39, CFI = .99, TLI = .99, RMSEA = .03, CI = .00 - .05, SRMR = .03, PCLOSE = .99), which provided support for the validity of the configural model and suggested that one model adequately represented data for both men and women.

Given the evidence of satisfactory model fit derived from the simultaneous analysis of multi-group data, measurement invariance across gender was assessed by constraining the factor loadings, factor variance, and the error covariance between Items 12 and 13 equal across men and women. Comparison of the fully-constrained measurement model ($\chi^2_{(47)}$ = 77.39) with the baseline model ($\chi^2_{(38)}$ = 52.63) resulted in a $\Delta\chi^2_{(9)}$ = 24.76 (p < .01), suggesting that the measurement model did not remain invariant across men and women. Sequential application of equality constraints to the factor loadings and examination of ΔB-S χ^2 indicated that Items 4, 6, 9, 12, and 13 were invariant across gender. However, Item 2 (“My worries overwhelm me”), Item 5 (“I know I should not worry about things, but I just cannot help it”), and Item 7 (“I am always worrying about something”) demonstrated variance across gender in that women displayed higher item means than men in the sample. Constraining the invariant factor loadings and the variance on the latent worry variable resulted in a $\Delta\chi^2_{(6)}$ = 10.11 (p = .12, *ns*), which indicated that only factor loadings on Items 2, 5, and 7 displayed significant group differences within the measurement model. However, it has been suggested that the ΔB-S χ^2 represents a stringent test of invariance (Byrne, 2010; Byrne, Stewart, & Lee, 2004), and the ΔCFI values for Items 2, 5, and 7 (.002, .003, and .005, respectively) fell at or below the .01 cutoff cited by Cheung and Rensvold (2002) as evidence of invariance. Consideration of the ΔB-S χ^2 values and the ΔCFI values together suggests that the factor loadings of the modified PSWQ-A demonstrated invariance across gender.

Additional invariance testing indicated that the error covariance between Items 12 and 13 and the variance on the latent worry variable remained invariant across gender, which provided further support for the invariance of the measurement model across men and women. Latent mean structure invariance tests were conducted wherein the invariant measurement parameters and the item intercepts were constrained equal, and the latent variable mean was constrained at zero for men. Comparison of the mean structure model ($\chi^2_{(51)}$ = 78.35) with the baseline multigroup model ($\chi^2_{(38)}$ = 52.63) resulted in a $\Delta\chi^2_{(13)}$ = 25.72 (p = .02), suggesting that the intercepts were not invariant across gender. Examination of the modification indexes indicated that the intercept associated with Item 6 (“When I am under pressure I worry a lot”) contributed to invariance across gender, with women displaying higher intercepts. Relaxing the constraint on the intercept for Item 6 resulted in significantly improved model fit in comparison to the baseline multigroup model ($\Delta\chi^2_{(12)}$ = 16.88, *p* = .15, *ns*). These results suggested that the modified PSWQ-A demonstrated partial scalar invariance. Examination of the latent group means indicated that women in this sample scored significantly higher (*p* < .00) on the latent dimension of worry than did men (unstandardized *M* difference = .30, *z* = 3.86)., suggesting that men and women demonstrated a significant group difference on the latent dimension of worry tapped by the abbreviated model of the PSWQ. The invariance tests for the modified PSWQ-A are summarized in Table 6.

Table 6

Invariance Tests Across Gender for the PSWQ-A

Model Description	B-S χ^2	*df*	ΔB-S χ^2	Δ*df*	CFI	ΔCFI
1) Baseline unconstrained multi-group model	52.63	38			.994	
2) Fully constrained model	77.39	47	24.76	9	.987	.007
3) Factor loadings and error covariance constrained equal	72.53	46	22.90**	8	.988	.006
4) Factor loadings constrained equal	71.54	45	18.92**	7	.988	.006
5) Factor loading 2 constrained equal	57.28	39	4.65*	1	.992	.002
6) Factor loading 4 constrained equal	53.20	39	.58	1	.994	.000
7) Factor loadings 4 and 5 constrained equal	60.32	40	7.69*	2	.991	.003
8) Factor loadings 4 and 6 constrained equal	53.38	40	.75	2	.993	.001
9) Factor loadings 4, 6, and 7 constrained equal	65.71	41	13.08**	3	.989	.005
10) Factor loadings 4, 6, and 9 constrained equal	60.28	41	7.66	3	.992	.002
11) Factor loadings 4, 6, 9, and 12 constrained equal	61.50	42	8.87	4	.991	.003
12) Invariant factor loadings and error covariance between Items 12 and 13 constrained equal	62.08	43	9.45	5	.992	.002
13) Invariant factor loadings, error covariance, and variance constrained equal	62.74	44	10.11	6	.992	.002
14) All invariant parameters and intercepts constrained equal	78.35	51	25.72*	13	.988	.006
15) All invariant parameters constrained equal; intercept for Item 6 relaxed	69.51	50	16.88	12	.991	.003

* *p* < .05, ** *p* < .0

Taken together, the results of the sequential invariance testing provided support for the invariance of the configural and measurement models of the modified PSWQ-A. The measurement model displayed partial measurement invariance as assessed by the most stringent difference tests but demonstrated consistent invariance across gender based on the ΔCFI values. The modified PSWQ-A displayed partial scalar invariance in that group differences emerged in terms of the intercept for Item 6. Group differences also occurred in the latent mean, with women scoring significantly higher on the latent worry variable.

Relationship of modified PSWQ-A to other meaures. Table 7 displays the correlations among the total scores from the full, 16-item PSWQ, the modified PSWQ-A, the BDI; and the BAI. The modified PSWQ-A demonstrated a strong positive correlation with the full-item PSWQ for men, women, and the total sample (all *r*s $\geq .95$). However, the correlation between the two measures was significantly higher for women than men ($z = -2.10$, $p < .02$). The modified PSWQ-A demonstrated moderate correlations with the BAI and BDI for men and women with no statistically significant difference across gender. However, men and women displayed a statistically significant difference in terms of the correlation between the BAI and BDI (r men = .55, r women = .71) in this sample ($z = 2.87$, $p < .01$).

Table 7

Correlations with Anxiety and Depression Measures for Men, Women, and Total Sample

	Men (*n* = 186)				Women (*n* = 316)				Total Sample (*N*=502)			
Measure	1.	2.	3.	4.	1.	2.	3.	4.	1.	2.	3.	4.
1. PSWQ		.95	.48	.61		.97	.58	.55		.97	.57	.58
2. Modified PSWQ-A			.47	.64			.57	.55			.56	.58
3. BAI				.55				.71				.68
4. BDI												

Note: All *p*s < .01 (2-tailed)

Discussion

The PSWQ assesses the frequency and intensity of worry, and the instrument represents the most widely- used measure of general trait in the extant literature. However, the extant literature has included a number of models for the PSWQ, and researchers (Brown, 2003; Brown et al., 1992; Carter et al., 2005; Fresco, Heimberg et al., 2002; Hazlett-Stevens et al., 2004) have not yet reached consensus about the underlying factor structure of the measure. Although a number of studies have compared the one- and two-factor models of the PSWQ, none of those studies have included comparisons of the abbreviated models of the measure that have emerged in the literature (Crittendon & Hopko, 2006; Hopko et al., 2003). This study utilized CFA to assess the relative goodness of fit of the different unifactorial and multifactorial models associated with the PSWQ. The seven *a priori* models tested in this study included five nested models derived from the full 16-item measure: the one-factor model conceptualized by Meyer et al. (1990), a one-factor model that included correlated errors (Brown, 2003); a two-factor model where positively- and negatively-worded items contributed to separate but correlated factors typically labeled as worry and absence of worry (Carter et al., 2005; Fresco, Frankel et al., 2002; Fresco, Heimberg et al., 2002; Hopko et al., 2003; Olatunji et al., 2007), and a three-factor model that included a general worry factor and two orthogonal method effect factors (Hazlett-Stevens et al., 2004; Pelleson et al., 2006). This study also examined the comparative fit of two non-nested models: an 11-item model that excluded five negatively-worded items and an abbreviated 8-item model that included only positively-worded items (Crittendon & Hopko, 2006; Hopko et al., 2003).

Given Brown's (2003) contention that a two-factor model allowed for error variance associated with the negatively-worded items, it was hypothesized that the two-factor model would provide better fit for the data than the 16-item, one-factor model. Furthermore, it was expected that the abbreviated 11-item and 8-item models would provide better fit than the 16-item, one- and two-factor models since these models eliminated the negatively-worded items. As expected, the results indicated that the two-factor model provided significantly improved fit compared to the one-factor model, which coincides with the results of other studies (Brown, 2003; Carter et al., 2005; Fresco, Frankel et al., 2002; Fresco, Heimberg et al., 2002; Olatunji et al., 2007). However, the fit indexes for the two-factor model that included separate but correlated worry and absence-of-worry factors did not meet the accepted criterion cutoffs indicative of optimal fit (Browne & Cudeck, 1993; Byrne, 2010; Hu & Bentler, 1999; MacCallum et al., 1996). These results suggest the two-factor model accounts for method effects associated with the negatively-worded items, but the model still contains points of misspecification. Brown's (2003) model with correlated errors (Model 2) provided better fit for the data, which supports his argument that both item wording and correlated error contribute to systematic method effects.

The one-factor, 11-item model assessed in this study demonstrated adequate rather than good fit for the data, whereas the abbreviated, 8-item model presented by Hopko and his colleagues (Crittendon & Hopko, 2006; Hopko et al., 2003) displayed the best fit among the six CFA models. It appears that the improved fit of the PSWQ-A (Crittendon & Hopko, 2006) stems from elimination of potential method effects associated with the five negatively-worded items and the content overlap between positively-worded items. Based on these findings and previous research that demonstrated that the absence-of-worry factor provided limited explanatory power in relationship to worry-related constructs, one can surmise that the underlying trait worry assessed by the PSWQ can best be conceptualized as a unitary construct.

The results of this study provide general support for the PSWQ-A model presented by Hopko et al. (2003) in that the abbreviated 8-item model provided the best fit for the data out of the seven tested *a priori* models. However, the RMSEA (.06) and PCLOSE (.12) values suggested that the model still included potential points of strain. Subsequent exploratory model modification suggested content overlap between Items 12 and 13, and adding error covariance between these items improved the model fit for the abbreviated one-factor model. Although the model in this study differed from the PSWQ-A in terms of the correlated error between Items 12 and 13, the abbreviated one-factor model appears to adequately represent both college students (this study and Crittendon & Hopko, 2006) and the clinical sample of older adults in the Hopko et al. study. Since Carter et al. (2005) found that the two-factor structure of the PSWQ did not hold across African-American and White-American college student samples and Pallesen (2006) found variance between community and student samples, the PSWQ-A (Crittendon & Hopko, 2006) may demonstrate less invariance across groups based on the data to date, but cross-validation studies with the abbreviated measure have not yet occurred.

Hopko et al. (2003) did not examine measurement invariance across gender in their clinical sample of older adults, but Brown (2003) found evidence of invariance across gender in the 16-item PSWQ using a clinical sample drawn from outpatient admissions at an anxiety disorder clinic. The results of this study are largely consistent with Brown's findings. Specifically, the measure demonstrated configural and measurement invariance across gender but displayed only partial scalar invariance across samples of men and women. Donaldson (2005) noted that different levels of invariance testing assess different levels of inference regarding equivalence across gender. The 8-item PSWQ demonstrated configural invariance, the most basic level of measurement invariance that indicates that the same items act as indicators on the same latent variable for both men and women (Chen, Sousa, & West, 2005). Invariance at the factor loading level suggests that the interval-level scaling of the observed variables or item indicators in the model are consistent across gender and that the average change in the latent variable per each unit change in the indicator is equivalent across groups (Chen et al., 2005, Donaldson, 2005). Without invariance at the factor loading level, one cannot assume that the two groups share a common metric and that the item indicators are free of construct bias. In other words, the items are not measuring the construct in the same way across groups. The abbreviated PSWQ displayed factor loading invariance based on the degree of change in the CFI statistic, but the more conservative change in χ^2 statistic implied that Items 2, 5, and 7 displayed variance across gender. Since change in CFI and change in χ^2 in the present study lead to different conclusion regarding the invariance of the factor loadings across gender, these findings must be cross-validated before assuming that the 8 item indicators in the abbreviated model assess the trait factor of worry the same way in men and women.

Factor loading invariance is a necessary and sufficient measurement condition for establishing a common metric between men and women, but both factor loading invariance and intercept or scalar invariance are required for what Donaldson (2005) referred to as "strong invariance" (p. 2346). Invariance in both factor loadings and intercepts establishes that the measurement regression lines are equivalent and changes in the latent variable across gender account for all systematic changes in the measured variable. The presence of strong invariance justifies inferences about true latent mean differences across groups (Donaldson, 2005). For the modified PSWQ, the scalar variance on Item 6 ("When I am under pressure I worry a lot") suggests that true latent mean differences between men and women could be at least partially confounded by the possible difference in the origin on Item 6. Since the findings of scalar invariance are at odds with Brown's (2003) results, it raises questions as to whether the PSWQ maintains measurement invariance across gender in clinical samples but not in nonclinical samples like those included in the current study. If invariance is present in clinical samples but not in nonclinical samples, it may mean that women in nonclinical samples view worry as more uncontrollable or intrusive in daily life than do men, but the differences in perceptions may disappear if men are experiencing a clinical degree of worry associated with anxiety disorders. Alternatively, men may be less likely to report subjective distress as defined by Items 2, 5, and 7 in that the content of these items seems to imply a lack of personal control. Acknowledging such feelings may run counter to conceptualization of perceived gender role for men, especially in a sample where the majority of individuals subscribe to a conservative faith.

Contributions

This study utilized CFA to assess the relative goodness of fit of the different unifactorial and multifactorial models associated with the PSWQ and addressed Hopko et al.'s (2003) call for independent research on the 8-item PSWQ-A (Crittendon & Hopko, 2006). Specific strengths of this study include a relatively large sample size comparable to other student and community samples (Brown, 2003; van der Heiden et al., 2010) and use of bootstrapped procedures to offset multivariate non-normality. Additionally, this study summarizes the growing body of research related to different PSWQ models. Only two studies have examined invariance across gender for the PSWQ in a clinical sample (Brown, 2003) and a Norwegian sample (Pallesen et al., 2006), and those studies produced mixed results in that Brown found invariance across gender but Pallesen did not. This study extended the available research by examining the sources of variance within the PSWQ-A. Although it is not clear whether the gender differences identified in this study are sample specific or extend to other groups, this study provides a jumping-off point in identifying the sources of gender variance in the measures.

Limitations

Although this study has contributed to the extant literature, several limitations should be noted. This study utilized a sample of college students, and it is unclear whether the results stem from characteristics specific to this sample or if the results will generalize to other student or community populations, especially since the sample was collected in northern Utah and 86% of the participants identified as LDS (Mormon). Additionally, the number of men in the present study was relatively small for CFA.

Another methodological issue related to the present study revolves around the adherence to fairly stringent criterion cutoffs established as indicators of good fit. Even though efforts were made to follow recommendation in the literature for goodness of fit, the strict criterion cutoffs across the CFI, TLI, RMSEA, and PCLOSE fit statistics prompted exploratory modeling to improve fit. If the less stringent criterion cutoffs established for adequate fit had been used in the present study, the correlated error term would not have been added to the PSWQ-A model identified by Hopko et al. (2003) and Crittendon and Hopko (2006). Although the correlated error term introduced in the present study appeared warranted, there is no question that the model modification reflects characteristics of this sample rather than a substantive model parameter. This work needs to be replicated and cross-validated to ensure that the findings in this study do not represent unwarranted capitalization on chance.

Future Research

The results of this study point t several possible venues for future research. For instance, Hopko et al. (2003) noted a strong correlation (*r* = .92) between the full PSWQ and the abbreviated measure. They suggested that the abbreviated instrument potentially represented an improved measure of trait worry in their clinical sample of older adults since it eliminated the difficulty of answering reverse-scored items and provided for quick screening of worry. However, the full, 16-item measure has previously demonstrated utility as a screening instrument for differentiating between GAD and no GAD; distinguishing between groups with GAD, GAD with comorbid disorders, and social phobia; and differentiating between cases and noncases of PTSD, social phobia, and depression (Behar, Alcaine, Zuellig, & Borkovec, 2003; Fresco, Mennin, Heimberg, & Turk, 2003). The results of this study have clinical implications in that dropping 50% of the original items from the PSWQ may fundamentally change the measure's specificity, sensitivity, and predictive power, thereby reducing the measure's utility as a screening measure. This is an area that warrants further research. Future studies could compare the screening utility of the full-scale and abbreviated models of the PSWQ.

The results of the invariance testing across gender need to be validated in other student samples. Such cross-validation studies could closely examine the relative contribution of the negatively-worded items by using models that specifically include method effects as a correlated latent variable, which would separate the variance into content and method or wording-related components (DiStefano & Motl, 2009). This would allow researchers to draw more definitive conclusions about whether the two-factor solution merely represents correlated method effects or whether the negatively-worded items represent a substantive factor. Understanding the extant of the method effects is a necessary step in addressing Brown's (2003) concerns about accepting models based solely on goodness of fit without consideration of interpretability.

References

American Psychiatric Association . (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (text revision). Washington, DC: Author.

Andrews, V. H., & Borkovec, T. D. (1988). The differential effects of inductions of worry, somatic anxiety, and depression on emotional experience. *Journal of Behavior Therapy and Experimental Psychiatry*, *19*(1), 21-26. doi: 10.1016/0005-7916(88)90006-7

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic* (2nd ed.). New York, NY: Guilford Press.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56,* 893-897. doi: 10.1037/0022-006X.56.6.893

Beck, J. G., Stanley, M. A., & Zebb, B. J. (1995). Psychometric properties of the Penn State Worry Questionnaire in older adults. *Journal of Clinical Geropsychology, 1*(1), 33 – 42.

Beck, A. T., & Steer, R. A. (1993). *Manual: Beck Anxiety Inventory.* San Antonio, TX: Psychological Corporation/Harcourt & Brace

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*(1), 77-100. doi: 10.1016/0272-7358(88)90050-5

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Behar, E., Alcaine, O., Zuellig, A. R., & Borkovec, T. D. (2003). Screening for generalized anxiety disorder using the Penn State Worry Questionnaire: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 34*(1), 25-43. doi: 10.1016/S0005-7916(03)00004-1

Blunch, N. J. (2008). *Introduction to structural equation modeling using SPSS and AMOS.* Thousand Oaks, CA: Sage Publications.

Borkovec, T. D. (1994). The nature, functions, and origins of worry. In G. C. L. Davey & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 5-33). New York, NY: John Wiley & Sons.

Borkovec, T. D., Robinson, E., Pruzinsky, T., & DePree, J. A. (1983). Preliminary exploration of worry: Some characteristics and processes. *Behaviour Research and Therapy, 21*(1), 9-16. doi: 10.1016/0005-7967(83)90121-3

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy, 41*(12), 1411-1426. doi: 10.1016/S0005-7967(03)00059-7

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour Research and Therapy, 30*(1), 33-37. doi: 10.1016/S0005-7967(92)90093-V

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B. M. (2010). *Structural equation modeling: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466. doi: 10.1037/0033-2909.105.3.456

Byrne, B. M., Stewart, S. M., Kennard, B. D., & Lee, P. W. H. (2007). The Beck Depression Inventory-II: Testing for measurement equivalence and factor mean differences across Hong Kong and American adolescents. *International Journal of Testing, 7*(3), 293-309. doi: 10.1080/15305050701438058

Byrne, B. M., Stewart, S. M., & Lee, P. W. H. (2004). Validating the Beck Depression Inventory-II for Hong Kong community adolescents. *International Journal of Testing, 4*(3), 199-216. doi: 10.1207/s.1532757ijt0403_1

Carter, M. M., Sbrocco, T., Miller, Jr., O., Suchday, S., Lewis, E. L., & Freedman, R. E. K. (2005). Factor structure, reliability, and validity of the Penn State Worry Questionnaire: Differences between African-American and White-American college students. *Anxiety Disorders, 19*(8), 827-843. doi: 10.1016/j.janxdis.2004.11.001

Castillo, C., Macrini, L., Cheniaux, E., & Landeira-Fernandez, J. (2010). Psychometric properties and latent structure of the Portuguese version of the Penn State Worry Questionnaire. *Spanish Journal of Psychology, 13*(1), 431-443.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*(3), 471-492. doi: 10.1207/S15328007SEM1203.7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. doi: 10.1207/S15328007SEM0902-5

Craske, M. G. (1999). *Anxiety disorders: Psychological approaches to theory and treatment.* Boulder, CO: Westview Press.

Craske, M. G., Rapee, R. M., Jackel, L., & Barlow, D. H. (1989). Qualitative dimensions of worry in DSM-III-R generalized anxiety disorder subjects and nonanxious controls. *Behaviour Research and Therapy, 27*(4), 397-402. doi: 10.1016/0005-7967(89)90010-7

Crittendon, J., & Hopko, D. R. (2006). Assessing worry in older and younger adults: Psychometric properties of an abbreviated Penn State Worry Questionnaire (PSWQ-A). *Anxiety Disorders, 20*, 1036-1054. doi: 10.1016/j-janxdis.2005.11.006

Davey, G. C. L. (1993). A comparison of three worry questionnaires. *Behaviour Research and Therapy, 31*(1), 51-56. doi: 10.1016/0005-7967(93)90042-S

Deffenbacher, J. L. (1980). Worry and emotionality in test anxiety. In I. G. Sarason (Ed.), *Test anxiety: Theory, research, and applications* (pp. 111-128). Hillsdale, NJ: Lawrence Erlbaum Associates.

Donaldson, G. W. (2005). Structural equation models for quality of life response shifts: Promises and pitfalls. *Quality of Life Research, 14*(10), 2345-2351. doi: 10.1007/s11136-005-3977-2

DiStefano, C., & Motl, R. W. (2009). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling, 16*(1), 134-146. doi: 10.1080/10705510802565403

Fortune, D. G., Richards, H. L., Griffiths, C. E. M., & Main, C. J. (2005). Worry and pathological worry in patients with psoriasis: Cross-sectional and longitudinal analyses of the Penn State Worry Questionnaire (PSWQ) in four samples of patients. *Journal of Clinical Psychology in Medical Settings, 12*(2), 143-152. doi: 10.1007/s10880-005-3274-9

Fresco, D. M., Frankel, A. N., Mennin, D. S., Turk, C. L., & Heimberg, R. G. (2002). Distinct and overlapping features of rumination and worry: The relationship of cognitive production to negative affective states. *Cognitive Therapy and Research, 26*(2), 179-188. doi: 10.1023/A:1014517718949

Fresco, D. M., Heimberg, R. G., Mennin, D. S., & Turk, C. L. (2002). Confirmatory factor analysis of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 40*(3), 313-323. doi: 10.1016/S0005-7967(00)00113-3

Fresco, D. M., Mennin, D. S., Heimberg, R. G., & Turk, C. L. (2003). Using the Penn State Worry Questionnaire to identify individuals with generalized anxiety disorder: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 34*(3-4), 283-291. doi: 10.1016/j.jbtep.2003.09.001

Gana, K., Martin, B., Canouet, M.-D., Trouillet, R., & Meloni, F. (2002). Factorial structure of a French version of the Penn State Worry Questionnaire. *European Journal of Psychological Assessment, 18*(2), 158-164. doi: 10.1027//1015-5759.18.2.158

Garson, G. D. (2008). Structural equation modeling. In *Statnotes: Topics in multivariate analyses.* Retrieved from http://faculty.chass.nesu.edu/garson/PA765/garson.htm

Hazlett-Stevens, H., Ullman, J. B., & Craske, M. G. (2004). Factor structure of the Penn State Worry Questionnaire: Examination of a method factor. *Assessment, 11*(4), 361-370. doi: 10.1177/1073191104269872

Holaway, M., Rodebaugh, T. L., & Heimberg, R. G. (2006). The epidemiology of worry and generalized anxiety disorder. In Graham C. L. Davey & Adrian Wells (Eds.), *Worry and it psychological disorders: Theory, assessment and treatment* (pp. 3-20). Chichester, England: John Wiley & Sons.

Hopko, D. R., Stanley, M. A., Reas, D. L., Wetherell, J. L., Beck, J. G., Novey, D. M., & Averill, P. M. (2003). Assessing worry in older adults: Confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychological Assessment*, *15*(2), 173-183. doi: 10.1037/1040=3590.15.2.173

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. doi: 10.1080/10705519909540118

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Lee, S. W., Stewart, S. M., Byrne, B. M., Wong, J. P. S., Ho, S. Y., Lee, P. W. H., & Lam, T. H. (2008). Factor structure of the Center for Epidemiological Studies Depression Scale in Hong Kong adolescents. *Journal of Personality Assessment*, *90*(2), 175-184. doi: 10.1080/00223890701845385

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, *20,* 975-978.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. doi: 10.1037/1082-989X.1.2.130

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology,* 70(4), 810-819. doi: 10.1037/0022-3514.70.4.810

Meyer, T. J., Miller, M. L., Metzger, R. L. & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28,* 487-495. doi: 10.1016/0005-7967(90)90135-6

Molina, S., & Borkovec, T. D. (1994). The Penn State Worry Questionnaire: Psychometric properties and associated characteristics. In G. C. L. Davey & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 265-283). New York, NY: John Wiley & Sons.

Morris, L. W., Davis, M.A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology, 73*(4), 541-555. doi: 10.1037/0022-0663.73.4.541

Olatunji, B. O., Schottenbauer, M. A., Rodriguez, B. F., Glass, C. R., & Arnkoff, D. B. (2007). The structure of worry: Relations between positive/negative personality characteristics and the Penn State Worry Questionnaire. *Journal of Anxiety Disorders, 21,* 540-553. doi: 10.1016/j.janxdis.2006.08.005

Pallesen, S., Nordhus, I. H., Carlstedt, B., Thayer, J. F., & Johnsen, T. B. (2006). A Norwegian adaptation of the Penn State Worry Questionnaire: Factor structure, reliability, validity, and norms. *Scandinavian Journal of Psychology, 47,* 281-291. doi: 10.1111/j.1467-9450.2006.00518

Papageorgiou, C. (2006). Worry and rumination: Styles of persistent negative thinking in anxiety and depression. In Graham C. L. Davey & Adrian Wells (Eds.). *Worry and its psychological disorders: Theory, assessment and treatment* (pp. 21-40). Chichester, England: John Wiley & Sons.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Stöber, J. (1998). Reliability and validity of two widely-used worry questionnaires: Self-report and self-peer convergence. *Personality and Individual Differences, 24*(6), 887-890. doi: 10.1016/S0191-8869(97)00232-8

Tallis, F., & de Silva, P. (1992). Worry and obsessional symptoms: A correlational analysis. *Behaviour Research and Therapy, 30*(2), 103-105. doi: 10.1016/0005-797(92)90132-Z

van der Heiden, C., Muris, P., Bos, A. E. R., & van der Molen, H. T. (2010). Factor structure of the Dutch version of the Penn State Worry Questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 204-309. doi: 10.1016/j-jbtep.2010.02.009

van Rijsoort, S., Emmelkamp, P., & Vervaeke, G. (1999). The Penn State Worry Questionnaire and the Worry Domains Questionnaire: Structure, reliability and validity. *Clinical Psychology and Psychotherapy, 6*(4), 297-307. doi: 10.1002/(SICI)1099-0879(199910)6:4<297::AID-CPP206>3.0.CO;2-E

Watts, F. N., Coyle, K., & East, M. P. (1994). The contribution of worry to insomnia. *British Journal of Clinical Psychology, 33,* 211-220.